Elementary Students' Comprehension of Computer Presenting Text

By

PAUL BIRD

B.Sc.(Hons.) Leeds University, 1972
Cert. Ed., Leeds University, 1973

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS

in
THE FACULTY OF GRADUATE STUDIES
(Department of Mathematics and Science Education)

We accept this thesis as conforming to the required standard

____________________________
Dr. Stuart Donn

____________________________
Dr. Lee Gunderson

THE UNIVERSITY OF BRITISH COLUMBIA
September 1990
© Paul Bird 1990
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Math and Science Education

The University of British Columbia
Vancouver, Canada

Date 25/9/90

DE-6 (2/88)
Abstract

The study investigated grade 6 students' comprehension of narrative text when presented on a computer and as printed words on paper. A set of comprehension tests were developed for three stories of varying length (382 words, 1047 words and 1933 words) using a skills hierarchy protocol. The text for each story was prepared for presentation on a Macintosh computer using a program written for the study and as print in the form of exact copies of the computer screen. Students from two grade 6 classes in a suburban elementary school were randomly assigned to read one of the stories in either print form or on the computer and subsequently completed a comprehension test as well as a questionnaire concerning attitude and personal information. The responses from the comprehension tests were evaluated by graduate students in Language Education. The data evolved from the tests and questionnaires were analysed to determine measures of test construct validity, inter-rater reliability, and any significant difference in the means of comprehension scores for the two experimental groups for each story.

The results indicated small but insignificant differences between the means of the three comprehension test scores for computer and print. A number of students reading from the computer complained of eye fatigue. The scores of subjects reading the longest story and complaining of eye fatigue were significantly lower.
# Table of Contents

Abstract.................................................................................................................. ii

Table of Contents................................................................................................. iii

List of Figures......................................................................................................... vii

List of Tables.......................................................................................................... viii

Chapter 1................................................................................................................ 1

Introduction ........................................................................................................... 1

1.1 The Purpose of the Study................................................................................. 1

1.2 The Need for the Study.................................................................................... 3

1.3 Overview of the Study...................................................................................... 3

1.3.1 Definition of Terms.................................................................................. 4

1.3.2 Null Hypotheses....................................................................................... 4

1.3.3 Research Variables................................................................................... 5

Chapter 2............................................................................................................... 6

Review of Literature............................................................................................... 6

2.1 Introduction....................................................................................................... 6

2.2 Reading............................................................................................................. 8

2.2.1 Reading Comprehension.......................................................................... 9

2.2.2 Reading Models....................................................................................... 10

2.2.3 Comprehension Assessment.................................................................... 14

2.3 Legibility......................................................................................................... 15

2.3.1 Legibility and Perception......................................................................... 15

2.3.2 Readability vs. Legibility......................................................................... 16

2.3.3 Legibility and Electronically Based Text.................................................. 17
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4 Typographic and Technical Factors</td>
<td>18</td>
</tr>
<tr>
<td>2.4.1 Type</td>
<td>19</td>
</tr>
<tr>
<td>2.4.1.1 Historical Perspective</td>
<td>20</td>
</tr>
<tr>
<td>2.4.1.2 Type Features and Legibility</td>
<td>22</td>
</tr>
<tr>
<td>2.4.1.3 Features of Typeface</td>
<td>23</td>
</tr>
<tr>
<td>2.4.2 Length of Line, Leading and Typographic Factors</td>
<td>25</td>
</tr>
<tr>
<td>2.4.3 Reading Surface, Colour, and Brightness</td>
<td>28</td>
</tr>
<tr>
<td>2.4.4 Speed of Reading of Electronically Based Text</td>
<td>29</td>
</tr>
<tr>
<td>2.4.5 Visual Fatigue of VDU Operators</td>
<td>30</td>
</tr>
<tr>
<td>2.5 Reading Comprehension and Computer Based Text</td>
<td>31</td>
</tr>
<tr>
<td>2.6 Conclusion</td>
<td>32</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>34</td>
</tr>
<tr>
<td>Procedural Methods</td>
<td>34</td>
</tr>
<tr>
<td>3.0 Introduction</td>
<td>34</td>
</tr>
<tr>
<td>3.1 Comprehension Test</td>
<td>34</td>
</tr>
<tr>
<td>3.1.1 Story Selection</td>
<td>35</td>
</tr>
<tr>
<td>3.1.2 Comprehension Test Construction</td>
<td>35</td>
</tr>
<tr>
<td>3.1.3 Collection of Data for Pilot Studies</td>
<td>38</td>
</tr>
<tr>
<td>3.1.4 Comprehension Test Scoring</td>
<td>38</td>
</tr>
<tr>
<td>3.1.5 Inter-rater Reliability</td>
<td>39</td>
</tr>
<tr>
<td>3.2 The Experimental Study</td>
<td>41</td>
</tr>
<tr>
<td>3.2.1 Experimental Population</td>
<td>41</td>
</tr>
<tr>
<td>3.2.2 Preparation and Presentation of Reading Material</td>
<td>42</td>
</tr>
<tr>
<td>3.2.3 Experimental Design and Procedure</td>
<td>43</td>
</tr>
<tr>
<td>3.2.4 Statistical Analysis</td>
<td>45</td>
</tr>
</tbody>
</table>
Chapter 4

Research Results

4.0 Introduction

4.1 Pilot Study 1

4.2 Pilot Study 2

4.3 Experimental Study

4.3.1 Summary of ANOVA Results

4.3.2 Correlation Results of Interrater Reliability

4.3.2 Measures of Normalcy of the Response Scores

4.3.4 Correlation of Comprehension Score and Reading Ability

4.3.5 Student Reactions to Reading Stories on the Computer

4.3.5.1 Responses to Question 12, Questionnaire B

4.3.5.2 Responses to Question 13, Questionnaire B

Chapter 5

5.1 Constraints of the Study

5.1.1 The Meaning of Comprehension

5.1.2 Research Population

5.1.3 Reactive Effect

5.1.4 Time Measurement

5.2 Discussion of Research Results

5.2.1 Pilot Study 1

5.2.2 Pilot Study 2

5.2.3 Experimental Study

5.2.3.1 Test Scores
5.2.3.2 Questionnaire Data................................................................. 63
5.4 Recommendations........................................................................... 65
5.5 Conclusions.................................................................................... 66
References............................................................................................ 68
Appendix A: The Program Written for the Study to Read Text Files......... 80
Appendix B: Samples of Print and EBT Used in the Study....................... 83
Appendix C: Copies of Comprehension Tests Used in the Study.............. 84
Appendix D: Questionnaires................................................................. 86
Appendix E: Responses to Selected Items on the Questionnaires.......... 88
List of Figures

Figure 1: Sign Definition in Peirce's Semiotic...........................................6

Figure 2: Z Scale Plot of Comprehension Scores for Freedom
   (Pilot 1)........................................................................................................45

Figure 3: Z Scale Plot of Comprehension Scores for Captive
   (Pilot 1)........................................................................................................45

Figure 4: Graph of Mean Scores for Comprehension Tests for
   Print and Computer Based Text.................................................................47

Figure 5: Z Scale Plot for Combined Computer and Print Scores
   for the Story Summer................................................................................50

Figure 6: Z Scale Plot for Combined Computer and Print Scores
   for the Story Captive................................................................................50

Figure 7: Z Scale Plot for Combined Computer and Print Scores
   for the Story Freedom..............................................................................51
List of Tables

Table 1: Measure of Variance for Comprehension Tests from Pilot Study 1.................................................................44

Table 2: Correlation Coefficient for Interrater Reliability of Pilot Study .................................................................46

Table 3: Anova of Test Scores for Captive with Print and EBT as Main Effect.................................................................48

Table 4: Anova of Test Scores for Summer with Print and EBT as Main Effect.................................................................48

Table 5: Anova of Test Scores for Freedom with Print and EBT as Main Effect.................................................................49

Table 6: Measures of Correlation Between Rater Scores for Experimental Study Comprehension Tests ....................49

Table 7: Coefficients of Correlation of Reading Ability and Total Scores for Experimental Study.............................51

Table 8: Summary of Response Data to Question 13.................................................................53

Table 9: Results of Anova of Test Scores and Reported Eye Discomfort.................................................................54
Chapter 1
Introduction

The invention of the printed word and moveable type in the 15th century brought language in written form to a larger audience than was previously possible with handwritten manuscript. The subsequent evolution of printed material has been a function of decreasing cost of materials and production, technical innovation of production methods and increased understanding of design elements to promote the author's intent and the reader's comprehension of the text. Print is a highly developed means of communicating ideas and information with culturally embedded reader expectations based on frequent exposure to the printed word in numerous forms.

However, electronically based text (EBT) is becoming an increasingly important means of handling the presentation, organisation and transfer of textual information. This is largely a result of technical innovation and decreasing cost and speed of processing information. Electronic text is now used extensively in word processing, data bases, accounting and electronic mail.

With the reduction in cost of computer hardware large numbers of North American school aged children have been exposed to computers as part of their everyday classroom experience. Major expectations of the use of microcomputers in this context have been to assist in the teaching of reading, as a source of textual information and for word processing (Lê 1989; Costanzo 1990). All these uses centre on the application of the computer to present written language albeit in different forms and with different reader expectations brought to the task of reading.
1.1 The Purpose of the Study

The growing number of computer based text applications aimed at children of school age has led to the question of whether presentation of text on a computer affects a reader's understanding of the material presented.

The computer screen differs in a number of ways from paper. The resolution of the image is much lower on a computer screen, the screen is a source rather than a reflector of light, and the amount of information on a computer screen is usually less than that of a page of paper.

The purpose of this study is to investigate the differences between grade 6 students' comprehension of continuous text as an electronic display and as print.

1.2 The Need for the Study

A number of studies (Muter et al. 1982; Gambrell et al. 1987; Haas and Hayes 1985; Feldmann and Fish 1987; Fish and Feldmann 1988) have reached the conclusion that there is no difference between comprehension of computer based text and print whilst another (Gerrell and Mason 1983) found evidence to the contrary. Thus despite considerable research the question of whether there is a substantative difference in the comprehension of computer based text and print is still an open one.

A common feature of these research efforts (with the exception of Haas and Hayes 1985) is the shortness of the text presented, usually less than 250 words, and the manner of evaluation, usually centred on the measurement of extraction of meaning from the text.
Reading short pieces of text may be considered a restricted exercise which does not reflect the overall nature of reading tasks as a whole. Evaluating text by assessing the comprehension of information embedded in the text assumes a narrow view of reading that does not take into account a reader's prior knowledge and understanding that she may bring to the reading task. This study takes into consideration the reading of longer pieces of text than has been considered in previous studies and means of evaluation that take into account the reader's role in constructing meaning.

1.3 Overview of the Study

Three narrative stories of varying length (382 words, 1047 words, and 1933 words) and appropriate reading difficulty for the grade 6 level were prepared for presentation on a computer and as print. Comprehension tests were constructed for each story. An experiment was carried out in which grade 6 students read these stories either on a computer or as printed words on paper. Students then completed a comprehension test and their reading comprehension assessed. Students also filled out questionnaires giving information of their personal computer use and their attitudes to the experience of reading the stories.

1.3.1 Definition of Terms

In this account the following terms are used:

*Electronically based text* (EBT) is used to describe any form of text that is presented through an electronic medium including computer bit mapped displays, video terminals, plasma screens. This study is concerned with computer based EBT.
Print is defined as text presented in conventional form as symbols on paper or an equivalent surface.

Informational text is defined as text that relates the particulars of an event or transaction.

Comprehension is defined as the ability to recall and/or make inferences from passages.

The three stories are referred to as Freedom, Captive, and Summer.

1.3.2 Null Hypotheses

The following null hypotheses are based on the premise that a quantitative assessment can be made of students' comprehension of narrative text:

1. Comprehension of narrative text displayed on a computer screen will be understood as well as the same text presented as print.

2. Comprehension will not be affected by prolonged reading of text on a computer screen.

1.3.3 Research Variables

The means of presentation, EBT and print, constituted the main independent variable of the study. Reading ability, gender, the presence of a computer at home, and previous reading of the story were also considered as subsidiary independent variables. The scores from the comprehension tests for the three stories constituted the main dependent variables of the study.
Chapter 2

Review of Literature

2.1 Introduction

The invention of the alphabetic writing system served as a means of encompassing language in a formalized structure. Olsen (1977 p.260) saw that this process "gave to Western culture many of its predominant features including an altered conception of language and an altered conception of rational man". Marshall McLuhan (1962) made a particular case for the revolutionizing effect that the printed word has had on the way mankind gained an understanding of himself and his environment.

The electronic media of computers and television have become an increasingly important means for man's interaction with the printed word. There is a growing integration of all forms of communication media which Negroponte (in Brand 1987) projects as being almost complete by the year 2000. Concomitant concerns have been voiced as to how electronic means of presentation will affect comprehension of information (Daniel 1985).

Many aspects of print have been thoroughly investigated by researchers such as Tinker (1963), Zachrisson (1965) and Spencer (1969). These studies, amongst others, have sought to investigate how the physical attributes of printed information affect the reading process. However, despite a long history of research the nature of processes involved with reading is still a speculative matter. Opposing views contribute to a debate which centers on whether meaning is embedded in the text or generated in the mind of the reader. The fundamental problem is still the one put forward by Johnson-Laird (1977 p.159) as "...simple to formulate: what happens when we read sentences?"
If the image on a computer screen is considered in semiotic terms as a component of a sign then de Saussure's (in Nadin 1988) definition of a sign can be considered. He promoted the idea that it could be seen as the unity between a signifier (the physical image) and the signified (what the image is meant to mean). Peirce (ibid.) advanced the idea that a sign could be said to be "something that stands for something in some respect or capacity" (ibid. p.47). From Peirce's definition the interaction of *representamen*, or that which represents, *object* or that which is represented, and *interpretant* or the process of interpretation (ibid. p.48) can be represented diagrammatically (Figure 1).

![Figure 1: Sign definition in Peirce's semiotic (adapted from Nadin 1988 p.47).](image)

It is the unity between the three components which represents a sign, and the sign can only be said to be constituted if the reader or observer establishes the relationship between the three. Thus in the case of a student reading (process of interpretation) from a page of text, the meaning (object) embodied in the text is represented by the printed word or images (representamen). This unity is disturbed if any component of
the triad is altered. By placing the object in a different context (or altering the representamen), as by displaying text on a computer, a different sign is constructed and the process of interpretation may subsequently be changed. There is the possibility that the interpretation of the object displayed on the screen may be fundamentally different to that of the object presented as print on paper.

This review looks at theories of reading and covers those aspects of the physical nature of print and electronically based text (EBT) that have been researched as possible influencing factors in the generation of meaning by the reader.

2.2 Reading

Reading is a complex activity which can be said to be one of the great cognitive achievements accomplished by developing children. It has been defined as the ability to extract meaning from any type of visual presentation (Duffy & Geschwind 1985) but in a narrower context represents an activity that involves extracting meaning from language encoded in alphanumerical or other symbols. Between 80 and 90 percent of normal children acquire this skill regardless of orthography or culture (Hirose & Halta 1988).

The extensive work of Tinker (1963) was directed toward studying the aspects of print that might affect the process of reading. Zachrisson (1965 p. 19) suggested that the reading process could be analyzed in terms of the following factors:

1. Reader. Prior knowledge, age, intelligence, reading experience, and physical equipment.

2. Text. Purpose and length.

4. **Situation.** Silent or oral reading. External conditions (such as lighting, posture, environment).

5. **Observation.** Attention, comprehension, reproduction, speed, fatigue, aesthetic evaluation.

This study is concerned with the question of whether there is a difference in the comprehension of text presented as print or on a computer. It can be argued that the computer alters the nature of presentation of text, the conditions under which text is read and the physiological response of the reader during the process of reading. These issues are discussed below.

### 2.2.1 Reading Comprehension

Reading comprehension can be seen to be both constructive and interactive (Flood 1984). The reader achieves understanding of what a writer has written by processing the visual information on the page and constructing meaning from the sensory input. Meaning is thus brought about by constructive processes. However, prior knowledge can be brought to the text by the reader which may influence how the text is processed. Information on the page and the reader's own knowledge interact to create meaning. The creation of meaning can thus be argued to be an interaction of the reader's knowledge and information embedded in text.

Various levels of information processing have been postulated and models constructed to explain observations of readers' behaviors. Levels associated with sensory input are seen as low order levels whilst those associated with semantic constructs are seen as higher order levels.
The reliance on models is problematic in that these are constructs that attempt to explain observable behaviors. These behaviors can take on the nature of established entities. Putnam (1981) refers to "metaphysical realism" to describe theories that have become engrained in the establishment thinking in the disciplines associated with cognitive science without sound reason. Lakoff (1988 p.150) suggests that "algorithms plus model theory are inappropriate for the characterization of meaning". He concludes with this analysis of the nature of how meaning is represented (ibid.):

Meaning is based on the understanding of experience. Truth is based on understanding and meaning. Innate sensory-motor mechanisms provide a structuring of experience at two levels: the basic level and the image-schematic level. Image-schematic concepts and basic level concepts for physical objects, actions, and states are understood directly in terms of the structuring of experience. Very general innate imaginative capacities can be put forward (for schematization, categorization, metaphor, and basic level physical concepts). Cognitive models are built up by these imaginative processes.

One of the assumptions of this study is that the basic sensory-motor mechanisms and subsequent construction of meaning proposed by Lakoff are affected by the formatting of text. The research giving credence to this assumption is outlined in Section 2.4 below.
2.2.2 Reading Models

Reading processes have been represented by functionally descriptive models which seek to explain how these processes take place. These can be generally divided into bottom-up, top-down and interactive models.

Bottom-Up Models

Bottom-up models dictate that information is processed at the lexical level. It may take place autonomously in a linear hierarchy (Forster 1979) and, in the case of Chomsky's model as interpreted by De Beaugrande (1981), relies on an automaticity and decomposition of syntax to generate meaning. Meaning is seen to be inherent in the syntactic structure and processing takes place in a serial and possibly a modular manner, that is, distinct from other processes. If the power of a model is a measure of its generalisability, the Chomsky model (as suggested by Smith 1985) has a low power due to the specific parameters that have to be defined in order to apply the model in any situation.

The process of reading becomes a serial input of information from which meaning is assembled. Gough (1972) indicated that reading involves a detailed linear motion that even registers the sequence of letters. Evidence for this is garnered from studies on the speed at which meaningful words and unrelated strings of letters are read out. This, it was suggested, gave indication that letters are first processed and then meaning is generated from the identification of words from the letter sequences.

Top-Down Models

This type of model actively involves the reader. Meaning does not reside primarily in the words of the text or in passages in isolation. The reader brings world views and
previous understanding to the reading task. The language provides a cue for the creation of meaning. Goodman (in Gunderson 1990 p.7) considers the process of reading to be "a psycholinguistic guessing game". Predictions of what the meaning of a word is come from higher level processes. As pointed out by Gunderson (1990) the connection between lower level sensory input and higher level cognitive processing is not provided for in this type of model.

**Interactive Models**

These models assume that meaning is constructed as an interactive product of text and context of various kinds which Spiro (1980) lists as including linguistic, prior knowledge, situational, attitudinal, and task contexts. As information is processed, the reader's understanding is constructed by these factors that are brought to bear on the reading task (Dennis et al. 1989). The resulting model of reading comprehension dictates that information from lower levels such as letter identification interacts with higher levels of processing such as syntactic analysis. Information flows from both the bottom up and the top down. The representations of understanding formed at each level are thus influenced by higher as well as lower levels. However Seidenberg and coworkers (1982) argued that certain low level processes are automatic and not accessed by other levels of processing. These are postulated to be independent of context. Conversely it has been found that poor readers are very heavily dependent on context for identifying words (Stanovitch 1980) which would suggest interactivity. Stanovitch's studies were carried out on isolating words in context. This may involve searching skills that do not necessarily represent the same activity as reading continuous prose.

The model suggested by McClelland (1986) is an attempt to represent an interactive perspective of comprehension. The underlying concepts of the model are as follows:
The processing system is organized into levels. These represent theoretical assumptions of how and where certain aspects of language processing take place. For reading these are a visual feature level, a letter level, a word level, a syntactic level, a word sense level, and a semantic or scenario level. At each level representations of meaning are generated.

The representation generated at each level is influenced by the activation of a number of simple processing units. The importance of this assumption is that representations are active and can be influenced by other representations at different levels.

Activation occurs through bi-directional interactions that occur both between levels and within levels. This assumes that the interactions are always reciprocal which makes the system interactive. The interactions within levels are competitive and bring about a dominant representation that is the 'best case scenario' for activating other levels. Although this model is more specific in outlining levels of sensory-motor input there is consensus that sensory-motor input is linked to higher level processes. The interaction of all levels, higher and lower, aid in the building of cognitive models.

The dichotomy of views on the subject of reading models is stated by Rudnicky and Kolers (1984 p. 249) as being the contrast between

...a view that holds that the physical signal is but the means of conveying language to the reader and that it remains only briefly in some short term store where its linguistic features are extracted and a view that holds that the physical signal is part of the mental representation of the words conveyed, an active participant in the processing.
A central question to this study is whether there is any qualitative difference between the stimulus of a computer screen and that of print which will affect the physical signal in a way to alter the comprehension of text.

2.2.3 Comprehension Assessment

Evaluation of a reader's comprehension of text read will depend on the theoretical stance taken by the evaluator. If the reader is seen as a processor of meaning provided by the text as provided for in bottom-up models then assessment of comprehension will depend on the ability of the reader to make sense of the writer's words. There is an emphasis on the 'correct' interpretation of the content. The validity of this construct will depend on the quality of the questions posed and how the responses are interpreted (Schwartz 1984).

Another view, explained by top-down models, considers the viewing of text as only part of the process by which meaning is constructed. Higher order cognitive processes are the prime factors that determine how text is understood. Thus the reader is the prime source of meaning and this meaning will depend on the prior knowledge, reading skills and reader attitude. Assessment of comprehension will focus on the meaning that the individual derives from the text.

A third perspective attempts to take into account both the meaning represented in the text and that brought to the text by the reader. This perspective is grounded in interactive theories of reading. Comprehension is constructed as the result of
interaction between reader and text (Spiro 1980). Meaning is not an object that has to be discovered from the text nor is it that which is brought to the text by the reader.

The English Assessment of Performance Unit (APU) Language Monitoring Team (in Thornton 1985) summarised the concerns underlying the development of their reading test materials. Among these was the need to make reading, even in the test situation, a meaningful activity. To that end reading material should be appropriate to the age-group in both difficulty and content.

Standardized reading tests have been criticized (e.g. Stierer 1985) for not measuring 'real' reading. This criticism is centred around the idea that the demands that a reading test puts on the student are "considerably different from those which the student encounters in and out of school" (ibid. p. 164). Stierer points out that reading tests are "blunt instruments which are only capable of measuring a limited range of reading skills" but to the user who is aware of these limitations the test can provide useful information (ibid. p. 166). Goodacre (1979) in his discussion of reading models makes the point that models of reading change. This change may render tests based on outdated models obsolete or lacking valid theoretical justification.

Most comprehension tests are considered by Schwartz (1984) to be attempts to judge how well subjects fulfill the expectation of the test constructor. Farr (in Schwartz p.62) referred to this as "thought getting".

The comprehension tests developed for this study utilise a 'real' reading activity akin to one that students might be expected to do as part of their school activities and for their own recreation. The responses were interpreted to allow for a student's own construction of meaning.
2.3 Legibility

2.3.1 Legibility and Perception

Much research has been carried out in attempting to isolate features of text presentation that present themselves as distinct variables (Tinker 1963; Zachrisson 1965; Gibson and Levin 1975). Work in the field of perception has generated evidence to support varying theories of information processing that support either serial or parallel and networking processing models. These stem largely from computational metaphors (Hulme 1984) the debate around which has been active since the earliest days of computer application to cognitive modelling (Gardner 1985) in which the issue of serial behaviour as a foundation to the predominant behaviourist stance of the time was questioned. The means of processing information is thus unclear at this stage. However there is ample research that has been undertaken to suggest the importance of textual features on the generation of meaning. This has been undertaken with two perspectives in mind. The first, to develop understanding as to the nature of the processes involved and the second, to develop textual arrangements that best suit the needs of the reader and enhance the generation of meaning.

2.3.2 Readability vs. Legibility

During the reading process there is, regardless of the model postulated, an initial interaction with the symbols on the printed page or the computer screen. Tinker (1963) made a distinction between the nature of readability, which was taken to be associated with the mental difficulty associated with the reading material and legibility which he considered to be concerned with the perceiving of letters and words and with the reading of continuous textual material. He defined it as dealing with 'the
coordination of those typographical factors inherent in letters and other symbols, words, and connected textual material which affect ease and speed of reading' (Tinker 1963 p.8). In this study legibility is defined in those terms.

2.3.3 Legibility and Electronically Based Text

Since the nature of electronically based text (EBT) differs from paper based text in a number of ways, new constructs of legibility have been postulated to encompass some of these differences.

Daniel (1985) sees that EBT has properties that may be distinct from conventional text. He proposes that conventional text can be considered in terms of static legibility. EBT on the other hand has both dynamic and interactive properties which will effect legibility. He referred to these aspects as dynamic and interactive legibility respectively.

Dynamic legibility refers to a "set of factors associated with how text and graphics can be moved relative to other factors and how this capability might affect reading performance" (Daniel and Reinking 1987 p.32). Since EBT is displayed transitionally as a function of an electromagnetic signal, the conditions under which it is displayed are flexible. Contrast, typographic layout and cueing can all be changed to suit the reader and the additional dimensions of animated illustrations and sound added to further contribute meaning (or confusion) to the text.

Interactive legibility is concerned with objects that can be linked to electronic text such as footnotes, dictionary, thesaurus and encyclopedic entries, and illustrations that may actually change what is read by individual readers (e.g. Weyer 1982, Duchastel 1986). The presence of unnecessary typographic cues has been shown to interfere with reading in good readers (Hartley and Davies 1976). With information available at the
behest of the reader, access can be made at the will. The depth to which a reader
searches for information in the text becomes a variable factor and the concepts of
legibility, as a descriptor of the physical nature of the text, and readability, as a
measure of the 'mental difficulty' of the text, become less clear.

2.4 Typographic and Technical Factors

The purpose of this study is to determine the effect that the typography and technical
factors of a computer display has on comprehension. Zachrisson (1965 p. 20)
considered these factors as relating to print in the following categories:

a. Type: (i) design, general shape; (ii) boldness, capitals; (iii) width (iv) capitals
   or lower case; (v) roman or italic.

b. Length of line.

c. Leading.

d. Color and contrast.

e. Typography: (i) spacing of words; (ii) margins and column distance; (iii) artistic
   presentation; (iv) arrangement of sentences.

f. Distance between text and reader or size of typeface.

g. Illumination.

The consideration of these elements of text presentation in the medium of the
computer screen has been a subject of study since the introduction of the cathode ray
tube (CRT) as a means of output display from a computer. In a comparison of
traditional page design and computer based information provided on-line Rubens
(1986) pointed out that with regard to any arrangement of textual and graphic material
the purpose is to provide an arrangement of that information to facilitate the user's task.

2.4.1 Type

Typefaces have been the subject of intensive creative input since their first appearance in the middle of the fifteenth century. They represent the fundamental unit of printed language. The nature of typeface is a complex issue that incorporates the subjective elements (design criteria) and the objective (perception). An original notion of typeface was that it had meaning embedded in the design. This matter is discussed in its historical context below.

2.4.1.1 Historical Perspective

The evolution of the alphabet represents one of the most important events in the history of mankind. The Western European alphabet has a history that appears to have originated with the Phoenicians, adapted by the Greeks (c. 11th century B.C.) and thence developed further by the Latins and Etruscans on the Italian peninsula (c. 9th century B.C.) (Gibson and Levin 1975). The refining of the Latin alphabet gave rise to the form found at the base of the Trajan column erected in 114 A.D. which Zachrisson (1965 p.171) considers the 'outstanding calligraphic achievement in the history of the alphabet' and the basis of a type design that has been in continuous use for nearly two thousand years.

Many different styles of typeface evolved some of which were designed to impart a special quality to the text such as a typeface reserved for royal decrees in 16th century France. At least one of these, Garamond is in widespread use today¹. This design,

¹ Adobe Corporation (Mountain View, California) introduced a new version of Garamond which was extensively researched from recasts of original Garamond typeface matrices in the collection of the Plantin-Moretus museum in Antwerp, Belgium. One of a series of redesigned classic typefaces, it is for use with Adobe's PostScript page description language for desk top publishing (McKinstry 1989).
attributed to Claude Garamond, has been widely copied by other type designers from the beginning of the 16th century. The italic font (1501), developed by the Venetian printer Aldus Manutius in 1501, reflected the inclined humanistic cursive calligraphy which was in use in renaissance Europe at that time. Aldus' intent was to develop an economical type that used less space and reflected his humanistic ideals. There was no intent to use this style in conjunction with the upright roman form. It was seen to be a vehicle for a representation of human ideas in a fashion more attune to cursive writing than the upright roman form. Capital italics were not developed since it was believed that capitals had a special meaning that was best represented by the roman form.

Peignot (1967 p.51) considered this a representation of 'le paradoxe de typographie'. If the letters are symbols by which the author's meaning is represented, to what extent should the form of the characters represent meaning *per se*? The persistence of Garamond as a typeface design raises the question of whether there is an innate quality to the design which leads to a perceptual advantage to the reader.

This paradox is in some ways reflected in the root of the word *text* which comes from the Greek *tekhne* meaning craft, and a more ancient Indo-European root meaning weaving. Bigelow sees this as an occidental association of letterforms with artificial or cultural products. He contrasts this with the oriental attitude to writing represented by Hsieh Ho (c. A.D. 500) and Qadi Ahmad (c. A.D. 1600) who represent letterforms as a "natural or transcendental phenomena" (Bigelow 1985). At one point Arabic writing became so cursive and ornamental that it became difficult to distinguish between letters, an obstacle which was overcome by the addition of diacritics and other features (Gibson and Levin 1975).
However, with the advent of infinitely variable fonts through programmed manipulation of typographical features (Knuth 1985), concern has been expressed about the cultural values embedded in typefaces and typography which Gürtler and Mengelt (1985) see as

...an expression of the human spirit, (which) are closely linked with the development of our society, ...they comprise values whose preservation and further development seem all the more urgently important to us when the technical developments of the means of production has practically no limits.

With the growing importance of text based information presented via computer screens in the last 30 years efforts have been made to improve the form of EBT type faces. Attempts such as the CMC 7 typeface developed for the European Computer Manufacturers Association were an effort to maximise the limitations of a low resolution C.R.T.. Peignot (1967 p.118) condemns this as 'une monstruosité typographique'.

As the resolutions of computer screens have increased so have the applications which use high resolution print output. Screen resolution is still not the same as that achieved by print however. The impression of a type face on the screen at 50 dots per square inch (dpi) for a low cost computer monitor to 120 dpi for a high resolution screen, that will only resemble a likeness to the output of between 300 and 600 dpi on paper from a laser printer. However, the effective resolution for typeset print is between 1200 and 2500 dpi (McKinstry 1989). With the use of two bit pixel screens the addition of shading to type has been shown to increase accuracy of the detection
of typographical errors on a computer screen to nearly 98% the rate as reading from paper (Brand 1987 p. 172). These are referred to as "fuzzy fonts" and have been integrated into the screen displays of the major computer manufacturers in the United States.

2.4.1.2 Type Features and Legibility

Tinker (1963) found that variation in the size of type and style of typeface had little effect on legibility within the range and size of typeface in common use. Italics were found to slow reading but no difference was found between boldface and ordinary print. Zachrisson (1965) found that a considerable range in typeface design and size could exist without significant effect on legibility. He was unable to detect significant differences in how roman and sans-serif letters were discerned.

Bouma (1978) defined the characteristics of letters as one of: a) acceptability or the degree with which the character fits the idea that the reader has of that character, b) identifiability or the degree with which the features of the letter stand out clearly and c) distinctiveness which is concerned with how different the letter or character is from others. In early attempts to generate clear typeface for EBT, Vartabedian (1971) found no difference in the time to complete a search task for text presented in different typefaces on a CRT. Gould and Grischkowsky (1986) found using 6 character sizes and 2 character sets displayed on a computer monitor that speed of reading was not affected within those sizes in normal use.

Hartley (1982) has demonstrated that research into readers' choice in typography (including typeface choice) is meaningless unless taken into consideration with other factors. His remarks were directed at Bell and Sullivan's (1981) results which he pointed out did not add anything new to the research on reader's preference to type.
Hartley and Rouum (1985) suggest that choices of type face are made on issues concerning nature of text, how it will be used and what typefaces are readily available. In using specific type faces to achieve effect there is evidence to suggest that readers associate the visual form of text with content. However Spencer suggests that congeniality (Zachrisson 1965) or appropriateness of typeface to a particular purpose is transitional. For instance, sans-serif types, which today are associated with 'modernity', were first revived in the eighteenth century because of "their association with rugged antiquity" (Spencer 1969 p.30).

2.4.1.3 Features of Typeface

Javal (in Huey 1968) illustrated how the leftmost part of English words carry more graphic information than the final parts and the bottom halves of letters carry more information than the top . Kolers (in Gunderson 1990) found for a person reading left to right the right hand side of the letter is more informative than the left. This result was confirmed by Gunderson (1990) in a study involving the isolation of letter cues.

Research carried out in the earlier part of this century sought to determine the relative legibility of letters as capitals and lower case letters. Spencer's review of this work shows that certain letters were found to be less legible than others in lower case but no difference was found between upper case letters (Spencer 1969). Vartabedian (1971) found that text presented in upper case on a VDU was searched 13% faster than that presented in lower case. This is in contradiction to Tinker's findings (1963) with print which showed that capitals were read 14-20% slower than lower case letters. However searching for discontinuous information and reading connected text are two different activities demonstrating different physiological responses in the reader (Bouma 1978).
Tinker found that sans-serif type faces were read as fast as serif types (Tinker 1963), a result confirmed by Poulton (in Spencer 1969). Prince on the other hand (in Watts and Nisbet 1974) found that serif letters in isolation or in nonsense words to be more legible than sans-serif but this result was reversed when the letters were arranged in identifiable words. Prince claimed this to indicate that the familiarity of the serif typeface influenced its legibility.

A study by Sanocki (1984) which investigated the cognitive representation of letter features suggested that visual information is represented in a structured system which becomes tuned in a specific way. This system tunes itself to different letters within the font but takes more time to adjust to different typefaces.

2.4.2 Length of Line, Leading and Typographic Factors

The length of line is dependent on the character size used. The commoner type sizes, 9 to 12 point, are all of about equal legibility. Very short lines appear to slow perception and increase the number and duration of eye fixations (Tinker 1963). The optimal line length seems to be one which accommodates ten to twelve words or 60 to 70 characters. Leading, or line spacing, allows for longer lines without loss of legibility (Spencer 1969). The average computer screen allows for this formatting of text. Earlier computers allowed only 40 characters per line, a condition which has all but completely disappeared from everyday computer use. Rudnicky and Kolers (1983) found that subjects read EBT more rapidly when text was displayed on a full or two thirds of a full width screen rather than when displayed on one third of a full width screen. EBT was also read more rapidly when set with 80 characters per line than 40.

Leading or line spacing can influence text structure and provide text cues (Hartley 1987). Tinker (1963) found leading to have an important affect on legibility of all type
sizes but less on 12 point type. Kolers et al. (1981) suggest that double spacing is preferable to single spacing when working with EBT and Wilkins (in Hartley 1987) considers that inappropriate line spacing may lead to fatigue when working with EBT. Empirical evidence for this was found in Lunn and Banks's (1986) study which sought to explain visual fatigue by the adaptation the eye has to make to the different means of leading found in EBT and print. In print, leading is measured from the imaginary line on which the characters are aligned relatively from one line to the next. It is usually set to create a line that is wider than the height of a character. In EBT, leading is most often measured in terms of the space that a line of characters would take up.

Researchers have also looked at the effect of justification or the parallel alignment of margins. Zachrisson (1965) found that in a group of subjects least proficient readers read unjustified text more quickly than justified text. Fabrizio et al. (in Spencer 1969) found no significant differences of speed and level of comprehension between subjects who read justified and unjustified text. However Trollip and Sales (1986) found that subjects read fill justified text on a computer slower than unjustified text. They sought to explain this result by the inconsistency of spacing that results when justification is applied on most word processing programs. They suggested that the spacing increases the number of eye regressions which are necessary in order to resolve the apparent inconsistencies of word spacing which arises from fill justification.

A number of considerations have been investigated with regard to the layout of text which can influence the reader's comprehension of text (Hartley 1987). In the current study only those factors which might be seen to be influenced by screen limitations are considered.

Duchastel considers the structure of text to be of semantic importance. In relation to computer based text he sees the important issue as “the interaction between a
person's cognitive structure on one hand and the embodiment of meaning within information technologies on the other" (Duchastel 1986 p.403). Structure is given by the rational ordering of information and the application of relevant layout techniques to reflect this. Work in this field (Meyer 1985, Thorndyke 1977) suggests that reading comprehension is affected by the typographic structure given to text.

Computer based information can be seen to differ from print based information in that the former can be semantically structured rather than format structured as in conventional text. This capacity of CBT is of great interest but is not the focus of this study.

Jonassen (1985) discusses the effects of a medium such as computer on learner performance by contrasting the notion of generative learning with mathemagenic control. The mathemagenic hypothesis (Jonassen 1985) suggests that certain design attributes of any medium give rise to learning. Cognitive structures are induced by the nature and form of the instructional material. The medium stimulates processes that would not be activated under normal circumstances. The generative model assumes that a learner when faced by stimuli, such as text books or computer screens, brings her own experience and attitudes to the task and constructs meaning from the event.

Typographic clues have the purpose of setting off important pieces of information to facilitate the retrieval of the information from memory, to distinguish supporting text from the main body, and to facilitate the reading of tables and list of figures (Glynn et al. 1985).

The two models discussed above seek to explain interaction with the text the form of which may be altered by the manipulation of these clues. Hartley and Davies (op. cit.) found the excessive inclusion of typographic clues confusing to readers, and Glynn and Di Vesta (1979) recommended their parsimonious use. However Tullis (1983) in
a review of factors involved in EBT determined these might be categorized into overall
density, local density, grouping, and layout complexity. He found some empirical
evidence that overall density, local density and grouping might have some affect on
human performance. The precise nature of this affect was not clear from the research
reviewed but there was no empirical evidence to suggest that increasing complexity of
layout interfered with human performance. He found this result to be in conflict with
design guidelines (e.g. Horton 1990).

As Hartley (1987 p.4) points out, the computer screen configures the screen in
"landscape" form, or wider than it is tall, rather than "portrait" form, or taller than it is
wide. This acts as a means of "chunking" text, that is, breaking it up into components.
Rubin (1986) considered reader expectation of the printed page and suggested that
any text layout on a computer screen should take into account those patterns
established with the printed page. Horton (1990 p.130) in the consideration of on-line
documentation suggests that although the "screen is not the page", screen based text
should resemble paper documentation with the proviso that certain text modifications
are necessary.

2.4.3 Reading Surface, Colour, and Brightness

The literature concerning the nature of the printed reading surface does not apply to
the computer environment since this research (See Tinker 1963) was concerned with
light reflected off paper and the computer screen is a source rather than a reflector of
light. Extensive work has been carried out, however, on the problems associated with
working at CRT workstations (Smith 1984; Reed & Williams 1988).

EBT has been displayed most often as light on a dark background though an
increasingly greater variety of display is available (e.g. Apple MacIntosh and graphic
user interfaces such as Microsoft Windows) which give negative contrast, dark on
white, as in conventional print. Negative contrast has been shown to result in fewer errors and reduced reading time in carrying out a task (Bauer and Cavonius 1980), better comprehension (Hume 1984), to be the preferred choice of users (Radl, 1980) and to have a marginal edge on legibility over positive contrast (Van Nes, 1986), though Cushman (1986) found there to be a significant difference.

Reynolds (1982) in a survey of display colour suggested that display should be governed by degree of contrast. Tinker (1965) made the point that legibility is not dependent on colour hues but more on brightness contrast. This in itself is distinct from colour contrast. Black on white was found to be the most legible whilst black on purple was found to be the least. Bruce and Foster (1982) carried out experiments on computer screens with 42 different combinations of coloured text and background in which the time taken for subjects to identify a line on a background was measured. The results tend to confirm Tinker's conclusions. Timmers et al. (1980) in experiments on the effect of decreasing contrast on the legibility of three letter words found a negative effect of decreased contrast on normal reading processes. They saw this as a result of shrinkage of the visual field with an increase in time required for word recognition. The implication of this research was seen to be applicable to the use of colour in displays which may involve varying degrees of contrast. In a comprehensive review of the role of colour in visual displays

2.4.4 Speed of Reading of Electronically Based Text

Because many researchers have found that reading is slower from a computer screen than paper, research has been carried out to determine variables of electronic display that might effect reading speed. Muter, Latremouille, Treurnier, & Beam (1981) found that reading from an electronic display took significantly longer than print but they used
a 48 cm low resolution colour monitor for the display so results may not be comparable
to studies carried out on computer displays which are usually of a higher resolution.

Gould and Grischowsky (1984) were unable to detect a change in proof reading
performance or vision acuity contrast between tasks displayed as EBT one day and
paper the next but subjects read 20-30% faster from print than off an electronic display.
Hsuah-Chih and Kam-Cheong (1988) measured reading efficiency which they defined
in terms of rate of reading and comprehension. Their results, which were concerned
with moving text on a screen found that when the movement of text was under the
control of the subject, reading efficiency increased with practice. Gould's research has
involved studies which have attempted to isolate variables influencing legibility of
screen (Gould et al. 1987). These have led him to move from considering what these
variables are to designing means of display from which people can read as fast as
from print. Indeed, he and his fellow workers have shown that when using an
advanced design of computer monitor reading can be as fast from a screen as from
print. His work was in conjunction with IBM and made use of a high resolution monitor
then under development.

2.4.5 Visual Fatigue of VDU Operators

The use of computer terminals and other VDU have been shown to cause detectable
temporary optical after-effects from tasks involving their use. This has been suggested
as a way of objectively measuring eye fatigue (Östberg, 1980). In a study carried out
by Haider et al. (1980), sustained VDU work was not shown to change visual receptor
functions but led to functional changes in temporary visual accommodation
mechanisms. As well, it was claimed that users exhibited prominent changes in their
states of well being. However Dainoff's (1980) study, although providing a
demonstration of a high incidence of reported visual complaints was unable to find an
objective optometric measure that correlated to these complaints. Nordqvist et al. (1986) in a study of reading text from videotex found no evidence of fatigue even with increased task difficulty. The slower reaction time in performance levels was taken to be an indication of a possible source of fatigue. Cushman (1986) sought to compare reader performance and ocular discomfort of subjects reading from microfiche, electronic display, and paper. Continuous text was presented in both positive and negative form. The results indicated that visual fatigue was not significantly greater for text presented in negative form as EBT than for paper. This contrasted to microfiche in which greater visual fatigue was found when reading negative images. Two types of common CRT display were used, with white (P4) and green (P31) phosphor screen coatings.

Lunn and Banks (1986) suggest that adaptation to the spatial frequency of lines of EBT could provide the reason for fatigue and the objective accounts of disturbed optometric measures.

A light phenomena with positive image on electronic displays was reported by Nylén (1985). It was perceived as a transient horizontal line on the screen. He suggested that it results from the eye moving at the same speed during the reading process as the electron beam moving across the display. The effect was dependent on the type of phosphor coating the display screen; the two types of phosphor in the experiment having different luminescent decay, it being most pronounced in screen coatings with a gradual rather than spiked decay characteristics.
2.5 Reading Comprehension and Computer Based Text.

The combined results of the investigations into the nature of screen based text has given rise to attempts to quantify any effect that the means of presentation of text on a computer will have on comprehension.

Gerrell and Mason (1983) found that fifth grade students comprehended computer based text better than printed material though Muter et al. (1981) in their study of text presented on television screens found no difference. Haas and Hayes (1985) found that graduate students were able to recall material more accurately from print than from a variety of computer displays, but the results were only statistically significant in the category of vertical location of information. Gambrell et al. (1987) in a study of comprehension and recall of EBT and print by grade 3 and grade 5 children found no statistical difference on measures of reading comprehension and concluded that extended reading of prose is a feasible proposition for young children. Feldmann and Fish (1987) found no difference in reading comprehension for following directions in groups of graduate students reading on and off computers and the same researchers (Fish and Feldmann 1988) found similar results in a study of elementary and high school students using informational and directional material. Both these studies were involved with short (90 to 250 words) pieces of writing that were related to normalized testing material and outdated achievement tests respectively.

2.6 Conclusion

From the research carried out in the last few years it is apparent that there is concern on the part of both academia and industry about the presentation of text on visual display terminals and computer screens. The nature of this concern is derived from both an interest in understanding the processes involved in reading and one of
developing the optimum conditions under which reading takes place. Both these lines of research are ones that have been under way since the beginning of the century in relation to print.

The continuation of research into the media of electronically based text (EBT) has led to the suggestion that there are different physiological reactions when reading EBT distinct from that of print. Although it has been hard to substantiate any precise effect, there is continuing concern and concomitant research into the ergonomics of working with VDU's.

Studies that have investigated the nature of the presentation and its effect on speed of reading lead to the conclusion that EBT is read more slowly than print by up to 25%. As the resolution of computer screens is increased and techniques are developed to enhance displays there is evidence that this difference can be eliminated. However most computer displays in use, especially in schools, will not benefit from these improvements in the near future.

Research into comprehension of EBT in comparison to paper has not shown any conclusive difference between the two media. The means of evaluation however has generally been limited to standardized texts and recall using short passages typically of about 250 words long.
Chapter 3
Procedural Methods

3.0 Introduction

The study fell into two components. The first involved designing and verifying tests of comprehension of three fictional narrative texts of varying lengths. The second involved an experimental study that sought to investigate the research hypothesis that there is a difference between comprehension of computer based narrative text and print. The null hypothesis, that there is no difference between comprehension of computer based narrative text and print and the subsidiary hypothesis, that this condition will be observed in increasingly longer pieces of text were tested by statistical analysis.

The means of presentation constituted the main independent variable of the study. The scores from the comprehension tests constituted the main dependent variables of the study.

3.1 Comprehension Test

The development of a suitable instrument to measure comprehension centered on two requirements of the study.

The first was dictated by the need to have a selection of narrative texts of varying length in order to test one of the central hypotheses of the study: that any disparity of comprehension of CBT and print would increase with length of text. This precluded the use of standardized instruments which tend to use shorter pieces of text, typically up to a maximum of 200 words. Most previous studies (e.g. Fish & Feldman 1988) have followed this direction.
The second was based on the need to develop an instrument that was centered on a consideration of comprehension as involving both meaning embedded in the text and that generated by the reader. This entailed the construction of instruments that were similar for each story and the concomitant testing of the instruments to assess their applicability in the experimental situation.

3.1.1 Story Selection

Three stories were selected from an inventory of short fictional stories used for prior reading studies based on their readability at the grade 6 level and length. Readability was assessed by the Flesch, Gunning-Fox and Flesch-Kincaid reading formulae (Harrison 1980). These values were gained by running the stories through a utility on the word processor Word.

The stories, Freedom (author unknown), Captive (author unknown) and All Summer in a Day (by Ray Bradbury) were 382, 1047 and 1933 words long respectively. A further factor for their selection was the likelihood of students having read the story before any assessment. The stories were given to 62 grade six and seven students in one Lower Mainland elementary school and 54 in another, none of whom reported having read the stories before. Although not conclusive, this was taken as evidence that the stories were sufficiently out of public circulation to allow a good chance that the experimental group would not have been exposed to them.

3.1.2 Comprehension Test Construction

Three comprehension tests were constructed by the author using a protocol developed by Hillocks and Ludlow (1984). This protocol is based on a model of the taxonomy of comprehension skills (Hillocks, 1980). The model has two main

---

1 Microsoft Corporation, Bellevue, Washington.
assumptions; namely that answers to questions represent skill types and a question
must be classified as a skill type in connection with the text from which it is derived.
The purpose of adhering to this model was to develop an equivalent set of questions
for each story.

Two types of questions were developed, literal and inferential. The rationale behind
this was the assumption that if readers cannot gain an understanding of literal
information they will not be able to draw inferences which may require that information.

The wording of each question was so determined that they contained no cue for
answering other questions in the set. They were submitted to experienced teachers
from the public school system and graduate students in Language Education at the
University of British Columbia, who tried to answer the questions using the information
contained in the set and also judged their suitability as means of assessing
comprehension of the text. Two questions from one set (Freedom) were changed as a
result.

Literal questions were constructed so that they could be answered with information
explicitly embedded in the text. They were designed to fall into three categories:

*Basic Stated Information*

The information for these questions is mentioned more than once, is essential for the
understanding of the story and is important for the higher levels of meanings of the
story.
**Key Detail**

Such information occurs at important junctures of the text and might bear causal relationships as to what happens in the story. The information appears less frequently than basic stated information.

**Stated Relationship**

The reader is asked to identify relationships that appear in the text. These are explicitly mentioned and are between two or more pieces of information such as characters, places or events.

Inferential questions required the reader to make generalizations about relationships that appeared in the text. They are of an appreciative and evaluative nature. The information did not appear directly in the text. They were divided into four categories:

**Simple Implied Relationship**

This is a similar category to the stated relationship with the difference that the relationship must be inferred. Such a question will rely on the reader's own experience and knowledge. Reference need be made to selected parts of the text.

**Complex Implied Relationship**

To answer this question the reader will require inferences based on a number of pieces of information.

**Author's Generalization**

This question attempts to tap the reflections that an author makes, consciously or unconsciously, on the nature of human beings and their environments.
Structural Generalization

To analyze structural generalization, a reader will have to be aware of how the author has put the story together to bring about an effect. It will require the reader to be able to generalize about the nature of the story and how effect is generated by that structure.

Seven questions were generated, one for each category. This was a departure from the methodology of Hillocks (1990) who developed two questions for each. This was justified by the brevity of the texts in comparison to Hillock's original investigation which used novels as sources of reading.

3.1.3 Collection of Data for Pilot Studies

Data were collected on the instruments in two separate elementary schools in the lower mainland involving 58 grade 6 students in one school (Pilot Study 1) and 24 in the other (Pilot Study 2) for the purpose of assessing the suitability of the stories for the grade level, developing rationales and criteria for assessment, and generating interrater reliability measures. Pilot Study 1 involved the testing of two of the three instruments to determine the suitability of the stories with regard to interest and readability. The third story could not be piloted at this stage due to time constraints. The comprehension tests were rated by the author.

Pilot Study 2 involved one class of students in French immersion who undertook the comprehension tests as part of the English language component of their program. All three instruments were assessed by two graduate students in Language Education. At this time agreement was reached between the raters as to how the answers were to be rated and general criteria developed. A rater reliability measure on initial agreement of scores was also calculated.
3.1.4 Comprehension Test Scoring

Following the protocol laid down by Hillocks and Ludlow (1984) criteria for "right", "partly right" and "wrong" answers were developed. Two graduate students in Language Education at the University of British Columbia undertook the scoring task. Two points were awarded for a right answer, one for a partly right answer and zero for a wrong answer. The criteria for correctness of response were developed by scoring a sample of responses from an urban grade six elementary class in Burnaby and evolving a consensus between the two graduate student raters and the researcher as to how responses could be evaluated.

Literal questions were scored with regard to how fully the responses could be seen to answer the questions. The questions were all answerable from information given in the text. Inferential questions required a response from which a logical relationship could be derived in the mind of the rater with the question. Thus for question categories 6 and 7 a wide variety of responses could be accepted that made sense in the context of the story.

3.1.5 Inter-rater Reliability

To improve consistency, raters marked all the responses for one question before moving onto the next and marked each set without discussion with the other rater.

Reliability for the study was computed by calculating the percentage of agreement of raters to the responses and the Pearson Product Moment Correlation Coefficients between the raters' total scores for each subject and the raters' total scores for each story question type. Pollitt and Hutchinson (1985) consider question scores to be a more valid test of correlation than subject scores.
The percentage agreement gave an absolute measure of the agreement whilst the Correlation Coefficients of scores gave a measure of how closely the scores of the raters correlated.

3.2 The Experimental Study

3.2.1 Experimental Population

The subjects for the study were selected from a suburban school in Burnaby, British Columbia. Students from two grade six classes participated with parental or guardian's permission. The resultant population resulted in a mixed ethnic group. Students in special programs, including those for English as a second language (ESL) and learning disabled (LD) students, participated in the study, however their scores were not included in analysis of the results. The ESL sub-population may not have had sufficient opportunity to develop reading skills in English and the LD group have individually demonstrated some dysfunction in learning to read. These populations are designated by procedures laid down by the Ministry of Education and involve the administration of prescribed standardized tests.

A total of 52 students participated, made up of 23 boys and 29 girls; 24 were from one class and 28 from the other.

Randomization of the normal reading population was achieved after blocking for gender in the following manner. Subjects were divided into two groups according to gender and then randomly assigned to one of the six experimental groups. Randomization was achieved by referral to a table of random digits (Hopkins and Glass 1984 p. 528). Blocking for gender was carried out to preempt any gender loading of the groups and skewing of results due to any gender specific effect.
3.2.2 Preparation and Presentation of Reading Material

For the experiment it was decided to use the highest resolution computer screen system commonly available. The reason for this choice was based on the growing movement toward higher resolution computer screens and the increased availability of these systems in elementary schools. If there is a difference in how information is comprehended from computer screens it is more relevant to evaluate the hypothesis on current hardware rather than outdated technology which is being phased out. For practical reasons this limited the choice to the Apple Macintosh Plus which has been implemented in large lab clusters in a number of lower mainland schools. This has a 22.5 cm diagonal, high resolution, 512-pixel by 342-pixel display.

A utility was prepared using compiled ZBasic\(^2\) which was designed to read the stories as text files on the computer in the form of a stand alone program. The route of using a compiled stand-alone utility was chosen over a file served utility such as Hypercard\(^3\) because of speed and reliability considerations when using Hypercard with a file served system. The screen layout consisted of two buttons, which activated forward and back pagination, and one which stopped the program. Subjects could only stop the program when they had come to the end of the story to increase the likelihood that a story would be read in its entirety before the student commenced answering the questions. Individual discs were prepared for each computer with system utilities sufficient to provide an independently bootable system.

The print material was prepared as an exact replica of the screen without the control buttons or screen menus. Typeface and font were exactly the same in both media. This was achieved using the Camera\(^4\) utility, a screen dumping program which made a

---

\(^3\) Apple Computer Inc., Cupertino, California
\(^4\) Source unknown.
facsimile copy of the screen and saved it as a file. The resulting files were arranged for printing using SuperPaint. Booklets were constructed so that the page size approximated the screen size. The resulting page of both book and computer screen allowed for twenty lines of text in 'landscape' configuration using 12 point New York type face.

3.2.3 Experimental Design and Procedure

A posttest-only randomized group two by three factorial design was adopted using two sets of three groups. Each group was randomly assigned to a particular story, the group being in turn randomly assigned to read either the print or computer version of the story. This arrangement can be diagrammed thus:

\[
\begin{align*}
SG_1 & \quad X_1 \quad O_1 \\
SG_2 & \quad X_2 \quad O_2 \\
SG_3 & \quad X_1 \quad O_3 \\
SG_4 & \quad X_2 \quad O_4 \\
SG_5 & \quad X_1 \quad O_5 \\
SG_6 & \quad X_2 \quad O_6
\end{align*}
\]

where \( SG_1 \) through \( SG_6 \) are subject groups who read the three different stories under reading conditions \( X_1 \) and \( X_2 \), computer and print, and \( O_1 \) through \( O_6 \) are the data from the results of the comprehension tests administered.

The experiment was conducted in a large classroom which houses the school's 28 Macintosh computers. It took place over two sessions during a morning, one session for each class. The computer screen luminosities were adjusted to an identical setting.
using a Minolta Flash Meter II calibrated at 100 ASA. Students were at liberty to change this if they so desired. All programs were set up ready to run from the desk top of the Macintosh operating system, a procedure that all students were familiar with as part of their regular use of the computer lab. Seating positions were allocated so that no adjacent subjects were reading the same story. The students reading the print versions of the stories were interspersed with those students reading text from the computers. The work stations for all students were thus physically identical with the exception that the computer groups' machines were turned on and the print groups' machines were turned off. All students commenced the task at the same time and all students in a story group, whether reading print or from a computer read and responded to questions on a paper worksheet. The subjects were timed in how long they took to complete the task by self reporting the time they finished. Access was allowed to the text during the completion of the comprehension test. Students also filled out a brief questionnaire giving information on previous computer use and their attitude toward the story and their reading of it.

Classroom teachers gave a measure of each subject's reading ability as derived from performance in classroom reading assessments. This was in the form of a grade (A, B, C+, C, C-, D) which was converted for the purpose of the experiment into a numerical rank equivalent.

Campbell and Stanley (1966 p.5) outline eight extraneous variables that might confound the effect of an experimental situation and subsequently bring the internal validity of the study into doubt. These are effects concerned with collecting experimental data over time and multiple testing. Because data were collected over a single morning many of these concerns (e.g. those related to extending an experiment over time) were not applicable to this design. One threat to validity would have been in the differences that might arise in the conditions between the two testing sessions.
This was checked for in an analysis of variance between the two scores achieved by the classes to check on whether there was a statistically significant difference. Another threat was the bias that might arise from differential selection of groups and one that relied on effective randomization to alleviate.

3.2.4 Statistical Analysis

The original protocol of Hillocks and Ludlow (1984) was centered on the development of a skills hierarchy. No effort was made to establish this hierarchy in this study and the subsequent item analysis using a Rasch model was not adopted due in part to the smallness of the sample (< 100) and evidence that classical inferential statistical methods are as informative as the Rasch model of analysis (Anderson 1990).

Rater reliability was established using the percentage agreement of rating for each response for each of the three stories and a Pearson product moment correlation of each of the subjects’ total scores as assigned by each rater. Correlations were also developed between the total scores for each category of question in the stories between the two raters.

For each story analysis of variance (ANOVA) was used to determine the significance of difference between final scores for the two experimental conditions. These were conducted taking into account gender, reading ability and the presence of a computer at home. These analyses were made to test the null hypothesis derived from the research hypothesis.

The three stories were treated separately since there was no sure guarantee that the comprehension tests were measuring the same aspects of comprehension despite the measures taken in creating similar questions for each of the instruments.
Correlations between subjects' reading ability as reported by the classroom teacher and performance on comprehension test scores were made using Pearson product moment coefficients.

The questionnaire response data for questions 12 and 13 on the student response form for those students reading from computers was categorized and subjected to statistical analysis. These ask the students for negative and positive comments on their experience in reading from the computer screen.
Chapter 4
Research Results

4.0 Introduction

The data were collected in three stages. The first set of data was collected in a pilot study (Pilot Study 1) to determine the applicability of the questions on the comprehension instruments for the grade 6 level, the second from a pilot study (Pilot Study 2) to establish rater reliability and to further assess the viability of the instruments with a different school population and the third from the experimental study to investigate the difference in comprehension of computer based text and print.

4.1 Pilot Study 1

The data from the initial pilot study were tabulated and entered into a Statsview II statistics program on an Apple MacIntosh SE computer. Measures of variance were calculated for the two comprehension tests evaluated (Table 1).

Table 1. Measures of Variance for Comprehension Tests from Pilot Study 1

<table>
<thead>
<tr>
<th></th>
<th>Freedom</th>
<th>Captive</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>57</td>
<td>31</td>
</tr>
<tr>
<td>Mean</td>
<td>5.1</td>
<td>8.0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>2.6</td>
<td>3.0</td>
</tr>
<tr>
<td>Variance</td>
<td>6.7</td>
<td>9.0</td>
</tr>
</tbody>
</table>

From these calculations Z scale plots were made of scores from the two comprehension tests to give indication of the normalcy of the distribution of the scores (Figures 2 and 3).
Figure 2. Z Scale Plot of Comprehension Scores for Freedom (Pilot 1)

Figure 3. Z Scale Plot of Comprehension Scores for Captive (Pilot 1)
4.2 Pilot Study 2

The data from the second pilot study consisted of scores from comprehension tests rated by the two raters enlisted for the study. Their scores for a sample of tests were tabulated and entered into the Statsview II statistics package on a Macintosh SE computer. Pearson Product Correlation Coefficients were calculated for the three studies to determine the correlations between total scores given by the raters for each subject and between the total sum of all the scores for each question as given by each rater. The results are tabulated in

Table 2. Correlation Coefficients for Interrater Reliability of Pilot Study 2

<table>
<thead>
<tr>
<th></th>
<th>Freedom</th>
<th>Captive</th>
<th>Summer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Students Total Scores</strong></td>
<td>.842 (N=10)</td>
<td>.862 (N=10)</td>
<td>.879 (N=6)</td>
</tr>
<tr>
<td><strong>Question Total Scores</strong></td>
<td>.886 (N=7)</td>
<td>.505 (N=7)</td>
<td>.966 (N=7)</td>
</tr>
</tbody>
</table>

4.3 Experimental Study

Data for the experimental study consisted of scores from comprehension tests from each rater, a final score derived from these scores and an arbitrator, information given by students as self reported data on home computer use, whether they had read the story before, positive and negative reactions to reading the story on the computer, and reading ability as given by the classroom teacher. No student reported having read the story before. The data were tabulated and entered into a text file and subjected to an analysis of variance (ANOVA) using the SPSS-X statistics package. The means of the different test scores for the two experimental conditions were
calculated. The results are displayed in Figure 4. Results for the ANOVA giving estimates for the significance of difference between the scores are tabulated in Tables 3, 4, and 5.

![Graph of Mean Scores for the Comprehension Tests for Print and Computer Based Text](image)

**Figure 4.** Graph of Mean Scores for the Comprehension Tests for Print and Computer Based Text
### 4.3.1 Summary of ANOVA Results

Table 3. Anova of Test Scores for Captive, with Print and EBT as Main Effect

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>Significance of F (a=.95)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Effect</strong></td>
<td>.982</td>
<td>.490</td>
</tr>
<tr>
<td>Computer at Home</td>
<td>.060</td>
<td>.811</td>
</tr>
<tr>
<td>Sex</td>
<td>.323</td>
<td>.581</td>
</tr>
<tr>
<td>Reading Ability</td>
<td>1.343</td>
<td>.316</td>
</tr>
</tbody>
</table>

Table 4. Anova of Test Scores for Summer, with Print and EBT as Main Effect

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>Significance of F (a=.95)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Effect</strong></td>
<td>1.008</td>
<td>.471</td>
</tr>
<tr>
<td>Computer at Home</td>
<td>.111</td>
<td>.746</td>
</tr>
<tr>
<td>Sex</td>
<td>.750</td>
<td>.407</td>
</tr>
<tr>
<td>Reading Ability</td>
<td>1.224</td>
<td>.360</td>
</tr>
</tbody>
</table>
Table 5. Anova of Test Scores for Freedom, with Print and EBT as Main Effect

<table>
<thead>
<tr>
<th>Main Effect</th>
<th>F</th>
<th>Significance of F (a=.95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer at Home</td>
<td>.084</td>
<td>.779</td>
</tr>
<tr>
<td>Sex</td>
<td>1.039</td>
<td>.102</td>
</tr>
<tr>
<td>Reading Ability</td>
<td>2.079</td>
<td>.156</td>
</tr>
</tbody>
</table>

4.3.2 Correlation Results of Interrater Reliability

Correlation coefficients were calculated as well as percent agreement between raters to determine interrater reliability for each story. Results are tabulated in Table 6.

Table 6. Measures of Correlation Between Rater Scores for the Experimental Study Comprehension Tests

<table>
<thead>
<tr>
<th></th>
<th>Freedom (N=15)</th>
<th>Captive (N=19)</th>
<th>Summer (N=18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ss Total Scores</td>
<td>.966</td>
<td>.935</td>
<td>.859</td>
</tr>
<tr>
<td>Question Total Scores</td>
<td>.899</td>
<td>.946</td>
<td>.958</td>
</tr>
<tr>
<td>% Interrater Agreement</td>
<td>85.7</td>
<td>79.6</td>
<td>80.7</td>
</tr>
</tbody>
</table>

4.3.2 Measures of Normalcy of the Response Scores

Z scale plots were calculated for the combined test condition scores of the three comprehension tests to give an indication of the normalcy of the distribution of the scores (Figures 5, 6, and 7).
Figure 5. Z Score Plot for Combined Computer and Print Scores for the Story Summer (N=18)

Figure 6. Z Score Plot for Combined Computer and Print Scores for the Story Captive (N=19)
The correlation between the subjects' final scores and reading ability as determined by classroom teacher was calculated (Table 7).

Table 7. Coefficients of Correlation of Reading Ability and Test Scores for the Three Stories Freedom, Captive and Summer

<table>
<thead>
<tr>
<th>Reading Ability and Test Scores</th>
<th>Freedom</th>
<th>Captive</th>
<th>Summer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.628</td>
<td>.353</td>
<td>.525</td>
</tr>
</tbody>
</table>
4.3.5 Student Reactions to Reading Stories on the Computer

Subjects filled out a questionnaire regarding their reactions to reading the stories. The questionnaire took different form for each reading condition (Questionnaire A for print groups and Questionnaire B for computer groups). Those subjects reading from the computer were asked the following two questions (on Questionnaire B as nos. 12 and 13):

12: What did you enjoy about using the computer to read?
13: What did you not like about reading the story from the computer screen?

4.3.5.1 Responses to Question 12, Questionnaire B

In response to question number 12 (What did you enjoy about using computers to read?) many commented on the novelty of the reading situation (e.g. #35 "it's just plain, more fun. (even if it is a boring story)"; #4 "I enjoyed clicking because when you read from a book you turn the page. Otherwise it was more less the same."), whilst others noted that they had something else to do other than reading (e.g. #41 "I enjoyed being able to do things with your hands while you read"). 4 subjects (#9, 22, 51, 58) stated that it was 'easy' or 'easier' (e.g. #58 "I enjoyed that it is just a lot easier using the computer"; #22 "I enjoyed reading from the computer because it is easy to read from"). Others remarked on the facility of using the computer (e.g. #53 "easy to find things, (pages words)"; #21 "That I could easily get to what page I want"; #5 "You can have an automatic page turner and you don't have to hold your book and save your spots").
4.3.5.2 Responses to Question 13, Questionnaire B

Responses to question number 13 (What did you not like about reading from the computers screen?) could be classified as those stating difficulties with the mouse (N=2), lack of illustrations (N=2), problems with the formatting (N=4), the screen being too bright (N=6), and “everything” (N=1). 9 respondents either gave a nil answer or did not respond (although all subjects responded to No. 12).

Having to use the mouse for page turning appeared as a drawback to some (e.g. #6 “had to click the mouse”; #1 “To keep on hitting the forward or back button”). The largest groups of responses centred around the formatting of the text and the brightness of the screen. The formatting of the text on the computer, although identical to the print format, was seen to be confusing (e.g. #4 “It’s easy to get lost in the screen”; #11 “you don’t know how many more pages to go...”; #33 “It’s hard to find the answers in it”; #41 “I did not like how it is the same all the time in looks”). Brightness of screen was reported as being uncomfortable (e.g. #26 “The light was to bright, I had to keep changing the color”)and in some cases, painful (e.g. #11 “...it’s to bright so it hearts your eyes”; #32 “My eyes started hurting and burning”; #51 “It hurts your eyes”).

Table 8. Summary of response data to question No.13

<table>
<thead>
<tr>
<th>Response Category</th>
<th>Subject #</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Everything</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Mouse Problems</td>
<td>1,6</td>
<td></td>
</tr>
<tr>
<td>Eye Discomfort</td>
<td>9,11,14,32,51,26</td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td>4,11,33,41</td>
<td></td>
</tr>
<tr>
<td>Lack of Pictures</td>
<td>42,48</td>
<td></td>
</tr>
<tr>
<td>Nothing Reported</td>
<td>22,21,5,17,38,53,58,34</td>
<td></td>
</tr>
</tbody>
</table>
The number of students complaining of eye discomfort for each story was related to the length of the story; Summer (1933 words) 3 complaints, Captive (1047 words) 2 complaints, Freedom (382 words) 1 complaints. Anova were calculated for the scores of Captive and Summer with eye discomfort as a factor. The results are tabulated in Table 9.

Table 9. Results of One Factor ANOVA of Test Scores and Reported Eye Discomfort

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>Significance of F (a=.95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer</td>
<td>.869</td>
<td>.032</td>
</tr>
<tr>
<td>Captive</td>
<td>3.651</td>
<td>.364</td>
</tr>
</tbody>
</table>
Chapter 5
Summary and Conclusions

5.1 Constraints of the Study

5.1.1 The Meaning of Comprehension

The purpose of this study was to investigate whether there is a difference between the way text is comprehended from a computer screen and print. The validity of the results rest largely on the instruments constructed and what they measure. These were constructed and evaluated to take into account interactive theories of reading. These theories suggest that reading is a constructive process that involves both the reader’s own knowledge and the information embedded in the text. From this standpoint there can be no wrong or right answers so much as those that reflect a greater or lesser relationship with information in the text, the extent of which is judged by the raters. A problem exists in that if the interactive paradigm is accepted there is no sure way of assessing comprehension that does not take into account the peculiar constructions that an individual may make out of the text. In the second pilot study one student interpreted the story Freedom as being about a taxi driver kidnapping the characters in the story. The student constructed a meaning that would appear to be heavily dependent on his own prior knowledge and less dependent on the text. From this a judgement is made by the raters that the student has not comprehended the story as well as the majority of the readers who saw the characters as being a blind person and her dog walking down a street. Is the kidnapping taxi driver interpretation a result of a fertile imagination and does this indicate impaired comprehension of the text? The means of evaluation is dependent on associating the observed behaviours of the students, that is their responses on the comprehension tests, with the concept of
reading comprehension. The evidence that the results show are in large part the outcome of educated guesses which have been structured through pilot studies and refinement of the raters' judgements of responses. As such they may show consistency but no claim can be made to relate that consistency to anything other than temporal understanding on the part of the raters as to what comprehension of the text means. The consistency at least gives some means of seeing whether there is a quantitative difference between the understanding of text on the computer and as print within the bounds of the meaning of comprehension developed by the participants in the study.

The fact that the test has a high inter-rater agreement is no indication that the test is a reliable measure of comprehension. The low to moderate correlation with teacher rating of reading ability is indication that the tests may be measuring an aspect of student performance that the classroom teachers are taking into account when assigning grades for reading.

5.1.2 Research Population

The study utilised the grade 6 population of a large elementary school. The population was thus limited (N=53) and may not represent a random sample of grade 6 students. The conclusions that can be drawn from the population may not be extendable into the general population with any degree of certainty. The small cell sizes of the experimental study meant that detailed statistical analyses of the experimental results could not be made.

5.1.3 Reactive Effect

A novelty effect for those using the computer may be present. The students participating in the study have been exposed to the Macintosh computers for two years
and used the computer room regularly (at least twice a week). However the task of reading a story off a computer was a new one and one on which all the students bar one who used the computer reported some positive aspect. This reactive effect to the different reading situation may alter the students' attitude and application to the task of reading and alter the conditions under which a valid comparative study can be made and conclusions drawn from it. One effect that might alter the generalizability of the study or its external validity is what Campbell and Stanley (1966 p. 6) refer to as the reactive effects of experimental arrangements. Subjects might react to the experimental situation in a different manner than a normal non-experimental setting. Since the computer based reading situation is essentially novel, an effective difference in reading comprehension may be disguised by a reactive effect due to the novelty of the situation compared to that of reading print and answering comprehension questions, a routine classroom activity. The population was chosen as one that has had extensive use of a variety of computer applications as one way to reduce this particular threat. However this threat could not be assessed to any degree of accuracy and remains as a caution in the final evaluation of the results.

5.1.4 Time Measurement

A valid and accurate measurement of time for reading the pieces of text was not obtained. Lacking this data limited conclusions can be drawn with regard to the possible deleterious effects of the physical signal of the computer screen on reading and the subsequent need to spend more time constructing meaning of the text.
5.2 Discussion of Research Results

5.2.1 Pilot Study 1

The first pilot study demonstrated the potential of the comprehension tests to measure an aspect of student's comprehension of the stories. Although a skills hierarchy model was used to develop the comprehension questions, no attempt was made to quantify the difficulty of the questions and to correlate a question type with difficulty. As explained in section 3.2.4 the item analysis carried out in the original validation of the hierarchical model put forward by Hillocks and Ludlow (1985) was not used. The z score plots gave an indication of the normal spread of the scores for the populations tested. These showed slight positive and negative skewness in the distribution. The questions for the story All Summer in a Day (Summer) were not tested in this initial study due to class time constraints. From this study it appeared that the questions being asked were relevant for the age group and resulted in a spread of scores close to normalcy. From this initial study it was also found that no student had read the story before and students found the stories interesting which gave credence to the suggestion that the stories represented a 'real' reading task (Thornton 1985).

5.2.2 Pilot Study 2

The purpose of the second pilot study was to develop a set of consistent marking criteria for the raters and to measure interrater reliability. The results indicated that a moderate rater reliability could be established. Time was spent by the raters discussing their rationale for marking and gaining further understanding of the text and the rationale behind the questions. A question on one of the tests was changed as a result of this process.
5.2.3 Experimental Study

5.2.3.1 Test Scores

A slight but consistent difference was shown between the scores for the two experimental conditions. However ANOVA results demonstrated no significant difference between the scores for the two experimental situations, EBT and print. No effect could be found for gender, the presence of a computer at home or reading ability (as determined by the class teacher).

From these results the null hypothesis could be accepted, i.e.:

\[ H_0: \mu_{ebt} = \mu_{print} \]

Where \( \mu_{ebt} \) is the mean of the scores for the students reading text on computers and \( \mu_{print} \) is the mean of scores for students reading print. The additional null hypothesis that there the difference between the means of the scores would not increase with length was accepted in the light of the acceptance of the primary null hypothesis.

Further analyses of interrater reliability gave increased correlations over the pilot study scores. The interrater subject score correlation was high (> .85) as was the interrater total question score correlation (> .89). Z score plots for the combined population gave indication that the questions asked were relevant to the student population and measured some aspect of understanding that was distributed across the population studied. There is no means of determining whether the understanding being measured is the same for each story though the following of a common protocol gave some consistency.

A measure of correlation of the students' test scores with their class grade in reading as determined by their class teachers gave a slight (.353) correlation for the story
Captive and moderate to substantial correlations for the other two stories, Freedom (.628) and Summer (.525). Any correlation might be taken as evidence of the construct validity of the test. Correlation would indicate that the test measures some aspects of reading ability that are taken into account by the classroom teachers when grading students for reading.

5.2.3.2 Questionnaire Data

The questionnaire gave information as to the negative aspects of reading off a computer. Six students complained of the brightness of the screen and discomfort of the eyes. Three of these students read the story Summer and demonstrated a significantly lower mean score as a group. The low numbers involved in the study do not allow any definite conclusions to be drawn from this result and further work is needed to clarify the situation. It might be suggested that the physical nature of the signal precludes the students concerned from concentrating on the text due to a sensitivity in the physical sensory apparatus to the peculiarities of a bright computer display. Another possibility is that the students represent a lower achieving group who have to spend more time reading text in order to make meaning of it and thus are likely to be subjected to eye strain due to the length of time needed to read the text. The luminance of the Macintosh computer screens used in the study were roughly twice that of the luminance of the paper used under identical lighting conditions. The question arises as to whether the brightness of the screen is presenting a perceptually different image or the physical sensory apparatus is unable to accommodate the brightness of the screen. Further work is needed to clarify this issue. The largest number of students to complain about brightness came from the group reading the longest story. Investigating the comprehension of longer pieces of text may be an area of fruitful study.
The confusion that one student detailed in terms of "getting lost in the screen" would indicate that there are some difficulties associated with the formatting that arises from the physical nature of the screen. If students are voicing difficulties associated with reading then these difficulties do not appear to influence the understanding of text that is being read. Again, the lack of an accurate measure of time taken to read the stories limits the conclusions that can be made as to what possible processes are taken place. The presentation of text on the computer was an exact copy of the corresponding printed page. As Daniel (1985) points out the computer screen presents a wholly different capability of text presentation in terms of legibility. His reference to dynamic vs. static legibility refers to the unique capacities of the computer to provide multiple dimensions to text. The concept of dynamic legibility refers to the capacity that computer based text may have to provide easily accessed multiple dimensions. Information can be quickly reached through the reader directing the search for specific information with minimum reading involved. The task of reading becomes different for each medium, print and EBT.

Using the semiotic construction outlined in Chapter 1, the evidence from this and other studies would suggest that the sign constructed on a computer screen differs significantly from that constructed as words on a piece of paper. The processes involved in interpreting the representamen (words on a computer screen) would appear to accomodate any difference between EBT and print insofar as the reader in most cases is able to construct meaning adequately. It is the presence of a section of the population, those who suffer discomfort in reading from the screen, which suggests that for some the sign is so different that their ability to construct meaning is significantly altered. A question is raised as to how else the differences between the two modes alter other aspects of the signs constructed to affect the construction of meaning. An example is the difference in the body position with which the reader
addresses EBT and print. Another is the dynamic aspect of EBT compared to print. The differences may or may not be negative in their outcomes as we perceive text and its uses at this time. What is important is to consider carefully the nature of the sign constructed, its intended purpose and whether the differences created by the computer derived medium is detracting from the intended purpose of the sign.

5.4 Recommendations

The growing importance attached to EBT suggests that students can expect to encounter increasing exposure to print displayed on a computer screen. The most likely source of text will be a student's own writing encountered during word processing activities. This involves a somewhat different reading process to the one investigated in this study in that the writer presumably has a good idea of the meaning of the writing.

Another common source of electronic text is in the form of encyclopedic and other database access. Again this involves a different reading task to one of reading narrative text. Meaning is likely to be more explicit and information concisely presented.

Both these tasks may involve short term focussing on the screen. With the task of writing, attention need not be fixed on the paper or screen for the cognitive processes involved with writing to take place. Reading from databases may involve precise short pieces of information though encyclopedic writing from CD-ROM sources may involve lengthy entries. Although no significant difference could be found between comprehension of EBT and print the study gave indication that some students may find reading from a computer screen to interfere with their ability to understand the text displayed. The effect may be one that increases with the length of text. Additional studies involving longer pieces of text may provide clarification of this issue.
With the limited narrative text currently available for display on computers discomfort associated with reading lengthy EBT may not be a significant problem at this time for school aged children. Gould et al. (1987b) concluded that a more fruitful pursuit than searching for reasons why there are problems associated with viewing computer screens is the development of screens that are effective and comfortable reading environments. This argument still begs the question that was raised in the first place as to whether there is a difference in the perceptual process of viewing a computer screen compared to paper and whether this is significant?

5.5 Conclusions

The study showed that there was a possibility of a difference in how text is read off the computer screen by some people. The small number of subjects in the experiment meant that the conclusions drawn could only be considered tentative. This study made every attempt to reduce the variables associated with typographic arrangements (e.g. typeface size and font, number of lines on a page or computer screen and line spacing). It appeared that most readers are able to successfully adapt their reading processes to the computer screen. There are a few who are not able to make this adaptation successfully and in semiotic terms the sign is altered sufficiently for them to lose sense of the object behind the text. It does not necessarily mean that the reading processes utilised are the same or different for print or paper. This is a matter for conjecture based on the observation of behavioural outcomes. We can only surmise as to the cognitive processes behind these outcomes.

This study is one in a number of studies conducted to address the specific issue of comprehension of computer based text. It is aimed at identifying and measuring an empirical effect. As discussed in Chapter 2 other studies have been conducted on the
speed of reading text on a computer and the conditions associated with optimal performance.

In other fields of computer use in education the field of word processing has been investigated (e.g. Daiute 1985) with regard to how writing becomes a different activity on the computer as compared to paper. Computers used in connection with measurement devices in school laboratories have been shown to bring about a different quality of understanding of experimental results through instantaneous transformation of data into useable forms (Driver 1990). This gives contiguous relationships between the scientific experiment and mathematical representation of the results in a form not previously available. The use of multiple dimension information linking in hypertext has led to different ways in which information can be stored and accessed (Garg 1988). All these, along with the studies done on reading comprehension, seek to explore the means of object representations offered by computers. These representations are varied and may be utilised through different and varied cognitive processes.

The search for empirical differences in performances although a useful endeavour may be misleading unless conducted in a wider perspective of understanding of how the signs, in semiotic terms, are constructed. As Gould et al. (1987b) suggested working on the machines to allow comparable performance between print and EBT so should work be carried out from a theoretical perspective as to why this should be so.
References


Appendix A

The following program, written in ZBasic™ for the Apple Macintosh™, reads a text file “README”. Two buttons, FORWARD and BACK load 20 lines of text at a time into the window. Another button QUIT stops the program and returns the computer to the operating system.

"This program sets up the standard Macintosh 22.5cm diagonal screen to read a text file "README"

BREAK ON

TEXT 2,12
DIM 255A$
DIM LocStore%(15)
DIM RecStore%(15)
Screen%=0
EndFlag%=0
OPEN "T",1,"TEXT:README"
WINDOW OFF :MENU OFF
COORDINATE WINDOW
GOSUB "Setup - Window #1"
GOSUB "Initial Read File"

"Main Loop"
DIALOG ON
A=DIALOG(0)
IF A<>0 THEN GOSUB "Event"
IF EndFlag%=1 THEN END
GOTO "Main Loop"

"Event"
B=DIALOG(A)
SELECT A
CASE 1: "Button selected - which one?"
SELECT B
CASE 1: "Back"
GOSUB "Retreat"
CASE 2: "Forward"
GOSUB "Advance"
CASE 3: "Quit"
GOSUB "Shutdown"
END SELECT
END SELECT
RETURN
“Setup - Window #1”  'Defines button locations
WINDOW #1, (0,0)-(525,345), 3
BX1=67:BY1=310:BUTTON #1,1,”Back”, (BX1,BY1) - (BX1+80,BY1+30), 1
BX4=361:BY4=310:BUTTON #2,1,”Forward”, (BX4,BY4) - (BX4+80,BY4+30), 1
BX3=214:BY3=310:BUTTON #3,0,”Quit”, (BX3,BY3) - (BX3+80,BY3+30), 1
RETURN

“Advance”  'Defines the “FORWARD” button
Screen%=Screen%+1
RecStore%(Screen%)=REC(1):LocStore%(Screen%)=LOC(1)
BUTTON #1,1
CLS
FOR X%=1 TO 16
LINE INPUT#1, A$
PRINT@(1,X%+2) A$
IF EOF(1) THEN GOSUB ”End Routine”:RETURN
NEXT X%
Count%=Count%+X%
RETURN

“End Routine”  'Defines the end routine
BUTTON #2,0  
BUTTON #3,1
PRINT
PRINT *  THE END*
RecStore%(Screen%)=REC(1):LocStore%(Screen%)=LOC(1)
Count%=Count%+X%
RETURN

“Retreat”  'Defines the “BACK” button
Screen%=Screen%-1
RECORD #1,RecStore%(Screen%),LocStore%(Screen%)
BUTTON #2,1
BUTTON #3,0
IF Screen%=0 THEN BUTTON #1,0
CLS
FOR X%=1 TO 16
LINE INPUT#1, A$
PRINT@(1,X%+2) A$
NEXT X%
Count%=Count%+X%
EndFlag%=0
RETURN

“Initial Read File”  'Counts the requisite number
BUTTON #1,0  
RECORD #1,0,0
RecStore%(Screen%)=REC(1):LocStore%(Screen%)=LOC(1)
FOR Count%=1 TO 16
LINE INPUT#1, A$
PRINT@(1,Count%+2) A$
NEXT Count%
RETURN

"Shutdown"
CLOSE #1
EndFlag% = 1
RETURN

' Returns program to desktop
you watched it. It was a nest of octopi, clustering up great arms of fleshlike weed, wavering, flowering in this brief spring. It was the colour of rubber and ash, this jungle, from many years without the sun. It was the colour of stones and white cheeses and ink, and it was the colour of the moon.

The children lay out, laughing, on the jungle mattress, and heard it sigh and squeak under them, resilient and alive. They ran among the trees, they slipped and fell, they pushed each other, they played hide and seek tag, but most of all they squinted at the sun until tears ran down their faces, they put their hands up to that yellowness and that amazing blueness and they breathed of the fresh, fresh air and listened to the silence which suspended them in a blessed sea of no sound and no motion. They looked at everything and savoured everything. Then, wildly, like animals escaped from their caves, they ran and ran in shouting circles. They ran for an hour and did not stop running.

ii. Print

you watched it. It was a nest of octopi, clustering up great arms of fleshlike weed, wavering, flowering in this brief spring. It was the colour of rubber and ash, this jungle, from many years without the sun. It was the colour of stones and white cheeses and ink, and it was the colour of the moon.

The children lay out, laughing, on the jungle mattress, and heard it sigh and squeak under them, resilient and alive. They ran among the trees, they slipped and fell, they pushed each other, they played hide and seek tag, but most of all they squinted at the sun until tears ran down their faces, they put their hands up to that yellowness and that amazing blueness and they breathed of the fresh, fresh air and listened to the silence which suspended them in a blessed sea of no sound and no motion. They looked at everything and savoured everything. Then, wildly, like animals escaped from their caves, they ran and ran in shouting circles. They ran for an hour and did not stop running.
Appendix C

Copies of the three comprehension tests used in the study.

All in a Summer Day

1. How often does the rain stop on Venus?

2. What did the children do to Margot before going outside?

3. How was Margot different from the other children?

4. When the boy pushed Margot what was she waiting for?

5. Why were Margot's parents taking her back?

6. What do you think the author is trying to tell us about how we treat other people? When you answer this question think about how Margot fits into the group and the reasons the children acted as they did.

7. The author does not tell us how Margot feels at the end of the story. In what way does he prepare us throughout the story so that we know how she might feel?

The Captive

1. Where is the captive imprisoned?

2. How did the captive's living nightmare begin?

3. Is it important to the guards whether the captive lives or not? How do we know this?

4. Why was the captive captured?
5. The captive is portrayed as despairing about his condition. Why do you think the captive's present condition is so painful to him?

6. What does the story lead us to think about how people treat prisoners such as the captive?

7. The author prepares us for a surprise ending to the piece. What details does the author provide us with to build up that surprise?

**Freedom**

1. What are Cathy and Trudy doing in the story?

2. Why is Cathy afraid?

3. In the situation in the story who is in charge of whom? How do we know?

4. What is Trudy's job?

5. Why do you think Cathy is nervous and does not trust Trudy totally?

6. Cathy is able to do things she was unable to do before. These are most likely things we would do without thinking. What do you think the story is trying to tell us about those things?

7. How does the story build us up for the last paragraph?
Appendix D

Questionnaires Completed by Students after Finishing the Comprehension Tests

Questionnaire A: Print Groups.

8. Have you read this story before?
9. Do you have a computer at home?
10. If you do have a computer at home what do you use it for mostly? (Circle one).
    Playing games
    Programming
    Writing
    Drawing
    Other:

11. Was the story easy to understand?
12. Do you enjoy using computers?
13. What did you not like about reading from the booklet?

Questionnaire B: Computer Groups.

8. Have you read this story before?
9. Do you have a computer at home?
10. If you do have a computer at home what do you use it for mostly? (Circle one).
    Playing games
    Programming
    Writing
    Drawing
    Other:

11. Was the story easy to understand?
12. What did you enjoy about using computers to read?
13. What did you not like about reading from the computer screen?
### Appendix E

Response Data for Questions 12 and 13 from Questionnaire B.

<table>
<thead>
<tr>
<th>Subject No.</th>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>Seemed to go faster</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>To keep on hitting the forward or back button.</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>It seemed to go faster then reading it on paper</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>nothing</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>I enjoyed clicking because when you read from a book you turn the page. Otherwise it was more less the same.</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>It's easy to get lost in the screen.</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>You can have an automatic page turner and you don't have to hold the book and save your spots.</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>There is nothing I don't enjoy about reading the story on the computer.</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>don't have to turn the page</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>had to click the mouse</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>It was easy to read the words.</td>
</tr>
<tr>
<td>9</td>
<td>13</td>
<td>It's hard reading of a screen because you're looking right into the screen.</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>you get to press the button on the mouse</td>
</tr>
</tbody>
</table>
you don't know how many more pages to go and it's too bright so it hurts your eyes.

It's quite fun

It hurts your eyes.

It was very interesting

Nothing really

That I could easily get to what page I wanted.

(no response)

I enjoyed reading from the computer because it is easy to read from.

Nothing.

I enjoyed just reading the story on the computer.

The light was too bright, I had to keep changing the color.

The ending

When they don't give us any clues

You don't have to hold anything you just have to press a button to go on to the next page.

My eyes started hurting and burning
If you want to go forward you just move the mouse to forward and then push.

It's hard to find the answers in it.

you don't have to hold on to anything

(no response)

it's just plain, more fun. (even if it is a boring story)

I liked everything

nothing

nothing

I enjoyed being able to do things with your hands while your reading

I did not like how it is the same all the time in looks.

you don't have to turn the page

no pictures

I enjoy reading it.

I didn't understand the beginning

It's different.

nothing

You can edjus how bright it is
48 13 it had no pictures

51 12 Easy
51 13 It hurts your eyes

53 12 easy to find things (pages, words)
53 13 nil

58 12 I enjoyed that it is just a lot easier using the computer
58 13 Nothing